EDUCATION ALIGNMENT

Joy Davidson (Digital Curation Centre, Humanities Advanced Technology and Information Institute, University of Glasgow)
Sheila Corrall (University of Sheffield)
George Coulbourne (Library of Congress)
Andreas Rauber (Vienna University of Technology)

Abstract

This essay reviews recent developments in embedding data management and curation skills into information technology, library and information science, and research-based postgraduate courses in various national contexts. The essay also investigates means of joining up formal education with professional development training opportunities more coherently. The potential for using professional internships as a means of improving communication and understanding between disciplines is also explored. A key aim of this essay is to identify what level of complementarity is needed across various disciplines to most effectively and efficiently support the entire data curation lifecycle.

Introduction

Over the last decade we have seen a vast increase in the general awareness about the need for data management, curation, and preservation activities. Indeed, many funding bodies are now seeking assurances at the bid stage from prospective recipient organizations and researchers that they are ready and able to manage access to their digital information over time. But just who is responsible for data management, curation, and preservation and how do those charged with responsibility get the skills they need to do the job? There is no single role within an organization that can take on the effective management of digital information from creation through to reuse in isolation. Researchers, information professionals, and information technologists all have roles to play. We need to determine how the various stakeholders can get the skills they need to effectively work together to undertake their

specific roles within the digital curation lifecycle.[1] While many of the essays in this volume look specifically at the topic of digital preservation, this essay will explore the broader concept of data management, curation, and preservation, which deals with data conceptualization and creation as well as preservation and access.

This essay will explore what educational alignment is needed—across disciplines as well as across nations—to support the digital curation lifecycle most effectively. In particular, the essay will focus on the skills of computing scientists, information science professionals, and researchers and how these may be progressed and supported through existing and emerging educational frameworks and knowledge transfer opportunities. There are views that data curation should be a profession in its own right as well as views that aspects of data management, curation, and preservation should be integrated to some degree into all disciplines.[2] To complicate matters further, we are beginning to see the emergence of new tools and applications that will significantly simplify, and in some cases automate, aspects of data management, curation, and preservation. As these tools are developed, the amount of specialist knowledge required by many of the stakeholders may be greatly reduced. As such, there is a real risk that some of the content covered in the new courses currently being developed may become immediately obsolete. Greater join-up between educators and those building infrastructure will be necessary.

This essay will look at where the community is now by highlighting some of the recent developments in embedding data management and curation skills in information technology, library and information science, and research-based postgraduate courses in various national contexts; move on to explore some of the challenges that we collectively face as well as some of the opportunities we might collectively exploit; and conclude with a series of recommendations for progressing the alignment of data management, curation, and preservation education within a number of disciplines. The authors would like to make clear that this essay

---

[1] See Digital Curation Centre (DCC) Curation Lifecycle model: http://www.dcc.ac.uk/resources/curation-lifecycle-model (last accessed 03-26-2012).

[2] It is important to note here that the authors did not cover education of the general public in our deliberations, however we recognize that this is a key area that requires investigation.

is not intended to represent a comprehensive investigation into the topic, but rather provides a snapshot of current national and disciplinary activities that may be of relevance as we attempt to improve the alignment of various curricula.

Where Are We Now?

Information Science

The UK has been at the forefront of developments in e-research as a result of its substantial investment in a national e-science program that has led to multinational collaborations and active participation by researchers at national, regional, and local levels. We have also seen significant national investment in research studies, development projects, and other initiatives intended to raise awareness, build understanding, develop policy, and enhance practice in the management of research data. The UK has the Digital Curation Centre, recognized internationally as a center of expertise and a catalyst for change, which has worked energetically and effectively with other key players in the UK and internationally, particularly the Higher Education Funding Councils' Joint Information Systems Committee (JISC), UKOLN, and the Research Information Network (RIN).

However, developments in the UK library and information science community have generally lagged behind other countries, notably the US. The tradition of library involvement in facilitating access to social science data is less well developed here than in North America. There is a small group of committed data librarians in the university sector who have responded positively to the current agenda[3] but we have not seen the creation of new positions, launch of initiatives, or development of services that has occurred in the US. Librarians have engaged in discussion through working groups and meetings initiated by Research Libraries UK and they have collaborated with others in exploring the need for a national research data service[4] but the lack of investment targeted specifically at university libraries combined with the funding cuts

[3] Macdonald, S. and Martinez, L. (2005) "Supporting local data users in the UK academic community," *Ariadne,* 44: http://www.ariadne.ac.uk/issue44/martinez/ (last accessed 03-26-2012).

[4] Lewis, M. (2010), "Libraries and the management of research data," in McKnight, S. (ed.), *Envisioning Future Academic Library Services: Initiatives, Ideas and Challenges,* pp. 145-168, Facet, London.

experienced by these libraries over the last few years has been a serious constraint.

While several universities across the EU have introduced digital preservation into their masters' level information science and library courses,[5] far fewer have started to include aspects of data management and curation. Part of the reason for this may be financial, with many schools of information and library studies experiencing cutbacks and lacking the funding for curriculum development that US counterparts have obtained from the Institute of Museum and Library Studies (IMLS) and National Science Foundation (NSF).[6] Current UK educational provision is limited and uneven, with minimal progress since the review conducted by Pryor and Donnelly in 2009.[7] Since then, UK-based library and information science educators have discussed the need for curriculum development at meetings of their subject association, the British Association for Information and Library Education and Research (BAILER), and have expressed interest in working collaboratively to meet needs for both initial professional education and continuing professional development (CPD). CPD provision is particularly important in the short term, but will be difficult to develop in the current UK situation unless additional funding is available.

Information Technology

To date, the majority of data management and curation research and development undertaken by information technology professionals has focused on preserving individual data objects. While some information technology programs do cover digital preservation topics—such as those offered by the Technical University of Vienna[8]—this is the exception rather than the rule.

---

[5] See e.g. the Information Management and Preservation (IMP) course at the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow http://www.gla.ac.uk/postgraduate/taught/informationmanagementpreservationdigitalarchivesrecordsmanagement/ (last accessed 03-26-2012).

[6] Ray, J. (2009) "Sharks, digital curation, and the education of information professionals," *Museum Management and Curatorship,* 24 94), 357-368.

[7] Pryor, G. and Donnelly, M. (2009) "Skilling up to do data: whose role, whose responsibility, whose career?," *International Journal of Digital Curation*, 4 (2),158-170: http://www.ijdc.net/index.php/ijdc/article/view/126 (last accessed 03-26-2012).

[8] Technical University of Vienna http://www.ifs.tuwien.ac.at/~andi/pr_thesis_topics.html (last accessed 03-26-2012).

What we need are information technology professionals who are capable of working collaboratively with other institutional stakeholders to develop scalable application chains and business processes that seamlessly integrate data management, curation, and preservation functionality over the entire lifecycle. Information technologists will need to be able to build bridges between disparate operational systems to ensure that data is generated and managed in a "preservation-ready" manner.

It is also vital that emerging information technology professionals are trained to think of the issues on a much grander scale. So far, we have been thinking in terms of gigabytes and terabytes, but new professionals need to be able to understand how error rates might scale as the amount of data that is produced on a daily basis increases exponentially.

At the moment, we are training information technologists to simply keep data alive through active intervention. We need to instigate a major shift in this way of thinking and emphasize the need for ongoing maintenance of end-to-end systems that are capable of producing and managing preservation-ready data. Essentially, there is a distinct need to produce information technology graduates who are fluent in enterprise architecture and interoperability. Indeed, these are areas that data curation and preservation practitioners are eager to progress and we need information technologists who can understand the problems and develop innovative approaches to meet these needs.

Data-centric Disciplines

In recent years, significant effort has been put into defining data management, curation and preservation roles and responsibilities for data authors, data scientists, and data mangers. In a 2005 report, the US National Science Foundation (NSF) defined data authors as "the scientists, educators, students, and others involved in research that produces digital data," and suggested that that data authors have a responsibility to:

- conform to community standards for recording data and metadata that adequately describe the context and quality of the data and help others find and use the data;
- allow free and open access to data consistent with accepted standards for proper attribution and credit, subject to fair opportunity to exploit the results of one's own research and

appropriate legal standards for protecting security, privacy and intellectual property rights;

- conform to community standards for the type, quality, and content of data, including associated metadata, for deposition in relevant data collections;
- meet the requirements for data management specified in grants, contracts, and cooperative agreements with funding agencies; and
- develop and continuously refine a data management plan that describes the intended duration and migration path of the data.[9]

Although these recommendations have been around for some time, there is little evidence that these skills are being embedded within UK postgraduate courses. Indeed, the data management skills and capacity session at the JISC Innovation Forum 2008[10] confirmed that while there were pockets of data management training within some UK universities' postgraduate courses, much more needed to be done to embed data management training into all postgraduate programs. To improve this situation, JISC recently funded the development of disciplinary-specific research data management training programs through their RDMTrain strand.[11] Researchers also have increasing access to a number of high quality support materials like those being produced by the UK Data Archive and the DCC[12] to assist them with their data management and curation activity.

Continuing Education

There has been a lot of work in recent years to develop intensive continuing education and training courses for data custodians and digital preservation practitioners, such as the

[9] See NSF report *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*: http://www.nsf.gov/pubs/2005/nsb0540/nsb0540_5.pdf (last accessed 03-26-2012).

[10] See JISC Innovation Forum 2008: http://jif08.jiscinvolve.org/wp/theme-2-the-challenges-of-research-data/sub-page-2/ (last accessed 03-26-2012).

[11] See JISC RDMTrain programme that is managed by the Managing Research Data Programme (MRD): http://www.jisc.ac.uk/fundingopportunities/funding_calls/2010/03/410dataskills.aspx (last accessed 03-26-2012).

[12] See the Digital Curation Centre (DCC) resources: http://www.dcc.ac.uk/resources (last accessed 03-26-2012).

Digital Preservation Management (DPM) Workshops, the Digital Preservation Training Programme (DPTP) that builds on the core concepts developed by DPM workshops, DigCCurr Institute, Digital Curation 101, and Digital Futures. A key objective of these courses is to bring together participants from a range of professional backgrounds to ensure that a wide variety of perspectives are shared and that viable curation approaches can be jointly developed and implemented at the institutional level. These courses have been quite successful to date and have led to some real changes in working practice within institutions.[13] However, the bulk of professional development training to date has focused on training those at the middle-management level with awareness-raising skills and the capacity to meet more immediate challenges. We also need to ensure that we target senior management staff and equip them to plan strategically over the longer term as well as furnishing those at the coal face with the practical skills they need on a daily basis.

Risks Presented by the Absence of Alignment on Core Competencies

The identification of basic data management skills for the various roles has been investigated by a number of ongoing working groups and initiatives (e.g., DigCCurrII, International Digital curation Education Action Group,[14] European MSc in Digital Curation working group, Knowledge Exchange, and DigCurV). Some progress has been made by these groups but the range of skills identified varies widely across these groups—ranging from very technical aspects to more traditional library and information sciences skills. To effectively embed data management and curation skills into a range of professions, agreement is still needed regarding: 1) what constitutes core data management and curation skills for each group; 2) a means of consistently describing and assessing these skills, and 3) a framework that supports the progression of skills development over time. Without agreement on core skill-sets and responsibilities for each of these groups along with an overall understanding of how they should all

[13] For example, the DigCCurr Institute and DC101 conduct post-course follow-up with participants and have learned about new activities that have been undertaken at their home institutions as a direct result of taking part in these courses.

[14] Hank, C. and Davidson, J. (2009) "International Data curation Education Action (IDEA) Working Group: a report from the second workshop of the IDEA," *D-Lib Magazine,* 15 (3/4): http://www.dlib.org/dlib/march09/hank/03hank.html (last accessed 03-26-2012).

fit together into the curation lifecycle we collectively face a number of risks. A selection of these risks are described below.

Relevance of Information Science Professionals

If education and professional development training in the library and information science sector do not evolve to cover data management and curation, there is a risk that librarians and other information specialists will not be able to contribute appropriately to the management of research data. Roles that library and information science staff are in other respects well qualified to fulfill may be assigned to other staff, who may have relevant and necessary subject and/or IT competencies, but lack the informational, managerial, and personal competencies that are required to apply the required specialist competencies successfully. The result could be that library services and information careers do not evolve in line with institutional needs and service provision is sub-optimal, being neither fit for purpose nor cost-effective. There could also be wider effects as stakeholders might lose confidence in the library and information science profession as a result of its perceived failure to respond appropriately to data management and curation needs and this might influence decisions on allocation of responsibilities and resources in other areas.

Capacity of Information Technology Professionals

Data management, curation, and preservation is not merely an evolution of "conventional" curation—complex and diverse as it may be—but rather an entirely new arena, requiring very different skills in completely new areas. This requires a completely new approach and new curricula to ensure we can educate information technology experts who will be able to develop and progress this field. There is a risk that we are not equipping information technologists with enough strategic development skills to produce appropriate systems to support researchers and information professionals in data management and curation activity. A solid foundational education and practical training opportunities are vital to ensure that information technology professionals can develop scalable, future-proof technical solutions and are also able to work with the tools and systems we already have in place to manage the less complex settings that we are currently confronted with.

Engagement of Researchers in the Curation Lifecycle

Managing and curating data requires dedicated effort. As such, researchers will want to ensure that their effort is being

allocated only to the items that they need to keep. Accordingly, researchers need to be educated to make effective selection and appraisal decisions about what they can keep and what they must not retain for legal reasons. The researchers themselves are best placed to determine just how much contextual information needs to be kept along with their data to enable validation of their research results, to provide evidence of good research practice, to meet their funding body and institutional requirements, and to facilitate reuse amongst their own community in the first instance.

Researchers have a key role to play during the conceptualization and creation stages of the curation lifecycle. Many of the decisions made during these early stages have major consequences for the curation and preservation of the data over the longer term. A particular area of risk for researchers is data licensing. Less ambiguity about what can—and more importantly cannot—be done with the data both in the short and longer term means that preservation actions may be more easily undertaken by preservation practitioners and restrictions on reuse are made clear to other researchers. If the data cannot be reused, there may be little value in curating and preserving it for the longer term.

Licensing, selection, and appraisal activities will involve several stakeholders in the curation lifecycle. Unless we can reach some agreement on specific data management roles, responsibilities, and (subsequently) skills, there is a real risk that we will duplicate effort across disciplines while failing to equip any of the stakeholder groups with anything but a thin veneer of knowledge across the lifecycle. Instead of educating the various groups on general issues we need to focus on specific areas where each group can make the best contribution and develop a deeper understanding of concrete actions and approaches. However, while there needs to be an emphasis on educating researchers in particular areas, it will also be essential that each stakeholder group understands how their own actions and those of the others fit into the bigger picture.

Development of Practitioners' Skills

There are many risks to be anticipated if data management, curation, and preservation skills are not covered by continuing education programs, field experiences, and professional internships. A major area of risk is in the ability to provide effective, market-ready graduate education in fields requiring management and curation of large digital collections, such as

library and information science (LIS), archival science, and museum studies. Within the LIS field, for example, researchers have noted gaps at times, e.g., between theory and practice[15] and between preservation studies, administration, and digital content.[16] These types of gaps must be bridged if tomorrow's corps of data management and curation experts is to be developed. In addition, the field has a need for professionals with expertise in advanced digital technologies, combined with other LIS expertise. In a 2006 study in librarians in academic and research libraries, participants identified digital topics (e.g., Web design, digital imaging, XML standards and technologies, and programming and scripting languages) as among the top areas in which their education had not prepared them adequately.[17]

A second major risk is that without optimal data management, curation, and preservation training, there will not be a cadre of experts able to provide support to users (including in libraries, museums, and archives) on digital resources in order to obtain maximum benefit. This is something that must be done as a daily service in addition to adequately planning and mapping the digital future for these collections and organizations. Thus, these needs are near-term and long-term. If collections cannot be managed, they will not be used and will lose value.

Another major risk is that of damage, deterioration, and/or loss of individual and shared heritage (or "memory") if individuals and organizations are not aware of the urgent need for data management and curation, are not able to support and conduct it, and if a workforce of data management and curation experts is not developed. One helpful analogy that has been drawn is between data management and curation preparedness and emergency preparedness.

---

[15] Ball MA. "Practicums and service learning in LIS education," *Journal of Education for Library* and Information Science. 2008;49(1):70-83.

[16] University of Michigan School of Information. *Engaging communities: fostering communities for preservation and digital curation.* Narrative by Elizabeth Yakel, Ph.D., of project proposal to the Institute of Museum and Library Services.

[17] Choi Y, Rasmussen E. "What is needed to educate future digital librarians: A study of current practice and staffing patterns in academic and research libraries," 2006. *D-Lib Magazine*, 12 (9): http://www.dlib.org/dlib/september06/choi/09choi.html (last accessed 5-21-2012).

Obstacles in Mitigating Key Risks

To mitigate the key risks described above, several obstacles need to be addressed in the short term. A selection of these is described below.

Limited Resources for Developing New Courses

Unsurprisingly, a lack of staff and/or financial resources can hinder efforts to embed data management and curation skills and activity into education, training, and the workplace. In the UK there have been substantial reductions in funding across the higher education sector as a result of the economic downturn. This general trend has been amplified for the LIS sector by government policy to concentrate resources on STEM subjects. Over the last two years, most information schools have experienced staff reductions through not being able replace departing staff, while student intakes at Masters and Doctoral levels have either stayed the same or increased. Many schools have also suffered cuts in the quality-related QR research funding[18] allocated on the basis of their performance in the Research Assessment Exercise because of the switch of resources from arts, humanities, and social sciences to STEM subjects. In addition, all academics face a continuing challenge in trying to balance the demands of research, teaching, and administration. There is constant pressure on faculty members to produce research publications and generate research income, which both tend to be prioritized over curriculum innovation—unless the latter is accompanied by significant external funding. Designing, developing, and delivering data management and curation curricula/courses that will meet identified education and training needs requires dedicated and sustained academic effort. Finding this academic staff time in light of these additional pressures can be difficult.

Modifications to Existing Curricula

Accommodating new content within existing educational curricula is a challenge, particularly in the core courses/modules/units taken by all students, where it is essential for data management and curation to be embedded so that all graduates gain the necessary knowledge and skills. There is pressure from employers and practitioners to extend or add

[18] For more information about quality-related research funding see: http://www.hefce.ac.uk/data/year/2011/quality-relatedfundingdata2011-12/name,67493,en.html (last accessed 3-27-2012)

coverage of specific topics to meet current needs (such as pedagogical knowledge and skills for development of information literacy), but rarely any suggestions of topics that could be dropped or given lower priority to make space for the desired new content. This problem is particularly acute in the UK, where the typical duration of a Masters programme is only one year and the taught element generally has to be squeezed into two 12-week semesters, with the remaining time devoted to an independent research project. The alternative strategy of embedding new competencies in specialist electives, pathways, or programmes (an individual course/module/unit or set of courses/modules/units) would in theory be easier to manage, but some UK schools do not currently offer electives and many others are under pressure to reduce the number of electives offered to students as a result of staffing cuts. Institutions would be unlikely to support new specialist programmes in the current economy—unless a critical mass of funded participants could be guaranteed.

Access to Practical, Hands-on Experience and Training

In supporting the growth of a skilled digital curation workforce, it is imperative that these professionals are able to activate the theory they learn in academic programs with practice. Internships and residencies address this key risk by tying theory with direct experience in the field. A number of challenges reside within the endeavor of planning and executing data management and curation internship programs, such as finding host organizations that are both actively practicing curation activity (as distinct from "digitization") and are geographically accessible to students; deciding where in academic programs it is best to incorporate internships (e.g., during the classroom experience or as a separate, freestanding component, such as a residency); establishing optimal length; determining who should lead planning and content (e.g., the home institution, the host, or guidelines from one or more professional bodies); and finding funds, apart from the currently small number of projects funded by agencies such as the US-based NSF, IMLS, and the National Endowment for the Humanities (NEH). The Laura Bush 21st Century Librarian Program of IMLS has funded a number of initiatives addressing the digital curation and electronic records needs of rural, underserved libraries. Other 21st Century projects focused on developing graduate certificates in digital curation and digital management and developing internship opportunities for the curation and stewardship of digital public information. These

projects support the need to address the future of libraries and the digital nature of their services. However, funding methods for field experiences in this area must evolve to include those other than just grants and government funding.

Among the components that will be important to include in field experiences are:

- Clear objectives;
- Short-term and long-term goals;
- Diverse skill sets, including in information technology; data creation (e.g., incentives, preservation, curation, and use agreements); data management (e.g., selection/evaluation, interoperable architecture, metadata standards, and maintenance); and data use (e.g., exploration, search/retrieval, authentication/verification, and use/reuse). Other desirable skills include policy, economics, project design, project and financial management, data rights/ownership, financial management, workforce development, and communications;
- An international component, e.g., through exchanges of professional data management and curation interns between the US and other regions such as Europe or Africa;
- Decisions on-site versus virtual components;
- One or more shareable final deliverables for interns, beyond the typical, brief exit summary—such as presenting at a meeting, writing a blog entry, or drafting a manuscript for journal submission;
- Means of assessing whether interns have acquired the skills during their internship;
- Providing community for interns, including through developing virtual communities and maintaining strong ties with home institutions; and
- Developing a training academy for internship supervisors and mentors, recognition in one or more ways for their efforts, and ways to build community.

The Office of Strategic Initiatives (OSI) at the Library of Congress has hosted interns for the past six years from several national internship programs and has seen the need for these principles firsthand. A major challenge in providing field

experiences is financial. In the Digital Preservation Outreach and Education (DPOE) Training Needs Assessment Survey[19] 34 percent of respondents said they did not have money for professional development or training. Studies have found that internships are highly advantageous—including to students, graduates, colleagues in the workplace, supervisors/mentors of interns, employers, academic institutions, and professionals—and that their return on investment is positive.[20,21] Internships do involve significant financial costs, but this is a challenge that can be addressed collaboratively; economic factors are a driver that can often bring people together.

There is also a lack of real-life, practical examples that can be located and easily repurposed for teaching and training to demonstrate the theoretical concepts taught about data management and curation in a meaningful and practical way. We need to be able to illustrate how data management relates to day-to-day activities in a range of disciplines. The work being carried out by the University of Michigan on graduate profiles and the DaMSSI career profiles[22] may be of value in addressing this challenge.

Supply and Demand

Most organizations are not devoting enough staff resources to data management and curation. In a 2010 national survey conducted by the Library of Congress's DPOE initiative, out of the almost 900 respondents, representing organizations of various types and sizes, only 33 percent had full-time or part-time paid staff dedicated to digital preservation duties. Eighty-four percent of respondents said, however, that their organizations consider it very important to preserve digital information for 10 years. The digital materials needing preservation included Web sites, architectural

---

[19] Library of Congress. Digital Preservation Outreach and Education (DPOE) Training Needs Assessment Survey: Executive Summary. 2010: http://digitalpreservation.gov/education/documents/DPOENeedsAssessmentSurveyExecutiveSummary.pdf (last accessed 03-26-2012).

[20] Sides CH, Mrvica A. *Internships: Theory and Practice.* Amityville, NY: Baywood Publishing; 2007.

[21] Lanier D, Henderson C. "Library residencies and internships as indicators of success: evidence from three programs," *Bulletin of the Medical Libraries Association*. 1999;87(2):192-199.

[22] University of Michigan graduate profiles, http://www.si.umich.edu/academics/pathways-success and DaMSSI career profiles, http://www.dcc.ac.uk/training/data-management-courses-and-training/career-profiles (last accessed 03-26-2012).

and design drawings, research data files, digital image files, PDFs, geographic information files, and audiovisual files. Challenges arising from the larger environment include the rapidly evolving digital landscape; issues related to shifts from analogue to digital; the uncertain economy and shrinking funding sources, the effects of which are being seen everywhere from local libraries to state archives to federal agencies; and the questions of where and for how long graduates trained in data management and curation can be employed, beyond the small number of projects based upon grants of limited duration.

This would seem to indicate that there is a market for graduates from a range of disciplines with data management and curation skills. However, recruiting sufficient participants to justify the time and effort required to develop new curricula/courses that include data management, curation, and preservation is a problem. The new student fees regime in the UK combined with the current depressed employment market is likely to affect recruitment at both undergraduate and postgraduate levels. Reductions in library budgets have affected both staffing levels and training support. The type of provision most likely to appeal to busy practitioners interested in CPD is flexible distance learning or an intensive immersion program (such as a short summer school). Both models could be costly to prepare and deliver, so providers would need to be confident that take-up would be sufficient for the fee income to cover their costs. The current financial situation is likely to deter potential providers from taking such initiatives—unless special funding were provided to cover the development costs. In the event that there is a clear demand for data curation and management training there is also a risk that the pool of trainers available to deliver courses will not be sufficient to meet this demand.

The recent *Riding the Wave* report[23] suggests that we should be educating "data scientists" and embedding aspects of both short-term data management and longer-term curation skills into all educational programs. However, can we guarantee that there is there a market for data scientists in disciplines outside of big science? Will students emerging with a specialism in data curation have enough knowledge in either the technical aspects of data management and the subject specific knowledge to be truly

[23] High Level Expert Group on Scientific Data *Riding the Wave*: How Europe can gain from the rising tide of scientific data report, October 2010: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf (last accessed 03-26-2012).

effective? The Spanning the Boundaries workshop[24] highlighted that employers still tend to hire from traditional backgrounds and train new staff on institution-specific approaches. So, will those emerging with more broadly scoped data management and curation qualifications be any better off?

Rapidly Shifting Goalposts in Information Technology

While analogue data curation itself has a strong tradition and a solid body of expertise to build upon and has been embedded into educational programs, the shift to digital is more than a mere change of representation or data carrier. As entire processes become digital, the ability to curate these will require the development of an entirely new body of professional knowledge, building on entirely new foundations that are tightly linked to the very basic differences of digital information and the way information is processed digitally. While listing all of these differences is virtually impossible, some of the crucial distinctions that need to be addressed in information technology education include:

Formats

The type of material to be processed is changing from static data to complex IT objects and entire systems that contain active code, transformation routines, and dependencies that reach all the way to specific details of the processing chain from a sensor via a range of IT systems with different hardware/software combinations and the technical implications of different computing paradigms such as the cloud, Web services and others, to interpretation in a closed loop. Information technology graduates will need a solid understanding of these dependencies on a fundamental technical level to be able to address any challenges arising from preserving data, processes, and systems.

Volume

The immense volume of digital objects, systems and processes to be managed requires a completely different level of automation. As a consequence, many steps that can currently be handled manually by experts will have to be

[24] Workshop run as part of iPRES 2010, see http://www.ifs.tuwien.ac.at/dp/ipres2010/workshops.html#ws1 (last accessed 03-26-2012).

automated, and thus ultimately implemented by technical systems if we want our solutions to scale. IT graduates will now need to be able to build scalable solutions rather than simply be able to develop short-term solutions. Equally, graduates must be capable of understanding security threats and developing means of protecting huge quantities of data.

Lack of Easy-to-Use Tools

The novelty of digital preservation at the level we are starting to see it now (i.e. preserving complex systems and processes rather than "just" PDFs or TIFF images) means that in many cases we do not even have available solid tool sets and procedures to train practitioners to manage them. We need to educate a large group of experts to help us develop the techniques that can then be used to train practitioners—while at the same time continuing to train practitioners to manage the less complex digital preservation settings that we have developed to date.

Lifecycle Management

Data management and curation will move from a largely ex-post activity, where an institution manages the data after its creation, to the operational systems, meaning that new IT infrastructure will need to be "preservation-ready." Preservation issues will need to be incorporated into the design and development process of such systems, rendering—in an ideal world—the actual curation invisible.

Opportunities for Alignment

There is no shortage of risks and challenges facing us as we attempt to better equip emerging graduates and existing professionals with up-to-date data management, curation, and preservation skills that they will need to perform well in their chosen professions. However, there are many opportunities that we can collectively seize to help overcome several of the challenges and risks described previously.

Defining Skill-sets and Facilitating Continuing Education

There is currently no consensus on the distinct range of skills, knowledge, behaviors, and attributes needed for different roles in data management and curation. However there have been several initiatives have been working to pin these down more concretely. Useful contributions to the discussion include Swan and Brown's

study[25] for JISC of *The Skills, Role and Career Structure of Data Scientists and Curators*; Pryor and Donnelly's mapping of core skills identified at the November 2008 DCC/RIN Research Data Management Forum against the four roles described by Swan and Brown, namely data creators, data scientists, data managers, and data librarians; and the outputs from relevant curriculum development projects, notably the IMLS-funded projects at the University of Illinois at Urbana-Champaign and the University of North Carolina at Chapel Hill, reported and discussed at a December 2008 workshop of the International Data curation Education Action (IDEA) Working Group.[26] Additional evidence can be found in published reports from library practitioners already involved in data management and curation.[27]

The DPOE initiative in the US has engaged in a review of the curricula of the main continuing-education data management and curation programs in the US and has collaborated with experts to establish a set of core principles for training. Among the skills required are skills in science, technology, engineering, and mathematics (STEM). There have been challenges in attracting students with STEM backgrounds to data management and curation education or careers, although efforts are being made in this direction. Skill sets from the humanities and social sciences are better represented. The RIN Information Handling Working Group[28] are also active in this area and are using the Vitae

---

[25] Swan, A. and Brown, S. (2008) *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs,* Report to the JISC, Truro: Key Perspectives http://www.jisc.ac.uk/publications/documents/dataskillscareersfinalreport.aspx (last accessed 03-26-2012).

[26] Hank, C. and Davidson, J. (2009) "International Data curation Education Action (IDEA) Working Group: a report from the second workshop of the IDEA," *D-Lib Magazine,* 15 (3/4) http://www.dlib.org/dlib/march09/hank/03hank.html (last accessed 03-26-2012).

[27] Gabridge, T. (2009) "The last mile: liaison roles in curating science and engineering research data," *Research Libraries Issues,* 265, 15-21 http://www.arl.org/bm~doc/rli-265-gabridge.pdf; Garritano, J.R. and Carlson, J.R. (2009) "A subject librarian's guide to collaborating on e-science projects," *Issues in Science and Technology Librarianship,* 57 http://www.istl.org/09-spring/refereed2.html#15; Henty, M. (2008) "Developing the capability and skills to support e-research," *Ariadne,* 55 http://www.ariadne.ac.uk/issue55/henty/; Witt, M. (2008) "Institutional repositories and research data curation in a distributed environment," *Library Trends,* 57 (2), 191-201 (all last accessed 03-26-2012).

[28] See RIN Information Handling Working Group: http://www.rin.ac.uk/resources/consultation-responses/joint-response-vitae-consultation-draft-researcher-development-fram (last accessed 03-26-2012).

Researcher Developer Framework (RDF)[29]—noted above—as a means of agreeing information handling skills among researchers.

Data management and curation has also evolved into a discipline in its own right, and as such it has created a community of experts from different backgrounds collaborating to tackle the challenges. With this evolution into a well-formed discipline, interdisciplinary working has become a good and well-lived practice—however potentially at the cost of becoming closed within its own interdisciplinary circles. We need to be aware that we need more external expertise by groups who do not necessarily see themselves at all related to data management, curation, and preservation activity and who currently are neither part of this community nor do they know of its existence. Examples, specifically within the IT domain, include hardware engineers, software engineering, distributed systems, algorithms, IT security, enterprise architectures, and many others—the contributions and cooperation of all of these are essential if we want to mitigate the data management and curation challenge from the onset, and to solve the challenges that need to be managed on a continuous basis.

While there are a number of data management and curation programs and professional development courses on offer, there is no easy way for prospective students to find, compare, and select courses that meet their immediate needs and allow them to plan for career development. Course offerings are usually self-contained with little if any reference to courses offered by other training providers. Without any means of contextualizing courses it is difficult to disambiguate and benchmark training options. Recent attempts to develop and/or refine competencies frameworks and to define specific skills-sets are described below.

Seven Pillars Model

The SCONUL (Society of College, National and University Libraries) Seven Pillars Model[30] helps to define a pathway from basic library and IT skills through to complete information literacy and describes progressive stages ranging

[29] See Vitae Researcher Developer Framework (RDF): http://www.vitae.ac.uk/policy-practice/165001/Consultation.html (last accessed 03-26-2012).

[30] See SCONUL Seven Pillars Model: http://www.sconul.ac.uk/groups/information_literacy/seven_pillars.html (last accessed 03-26-2012).

from the novice to the expert. While the model has proved valuable as a planning tool among UK HEIs, the model developers felt that the model would benefit from additional details in data management aspects. A revision of the model was undertaken in during the first half of 2010 to incorporate additional data management elements.[31]

Researcher Development Framework (RDF)

In November 2009, Vitae released a draft of its Researcher Development Framework (RDF) for public consultation. The RDF is intended to be used as a "tool for planning, promoting and supporting the personal, professional and career development of researchers. It describes knowledge, skills, behaviours and personal qualities acquired by researchers and encourages researchers to aspire to excellence through development to a high level."[32] The RDF offers great potential for describing basic data management skills required at each stage of a researchers' career and for securing agreement on basic skill sets. However, while data management skills were implied throughout several sections of the draft RDF, they were not as explicit as they perhaps should have been. In November 2009, the Research Information Network (RIN) established an information handling working group[33] which developed a response to the draft RDF that included recommendations for more explicit data management skills at each of the RDF stages. Many of the WG's recommendations have since been incorporated into the revised RDF.[34] The working group has remained active and is working to improve the provision of information literacy education within UK HEI programmes. The working group includes members from relevant bodies including the Society for College, National and University Libraries (SCONUL), Research Libraries UK (RLUK), the Chartered

[31] See SCONUL Seven Pillars Model: Research Lens Model table of skills and attributes: http://www.sconul.ac.uk/groups/information_literacy/sp/researchtable.jpg (last accessed 03-26-2012).

[32] See RIN Information Handling Working Group response to Vitae RDF: http://www.rin.ac.uk/resources/consultation-responses/joint-response-vitae-consultation-draft-researcher-development-fram (last accessed 03-26-2012).

[33] See information about the working group: http://www.rin.ac.uk/mind-skills-gap (last accessed 03-26-2012).

[34] See Vitae Researcher Development Framework (RDF): http://www.vitae.ac.uk/rdf (last accessed 03-26-2012).

Institute of Library and Information Professionals (CILIP), the Digital Curation Centre (DCC), the British Association for Information and Library Education and Research (BAILER), the Higher Education Academy (HEA), the UK Council for Graduate Education (UKCGE), and the Joint Information Systems Committee (JISC).

DaMSSI

Led by the Digital Curation Centre (DCC), this Data Management Skills Support Initiative (DaMSSI) aimed to facilitate the use of tools like SCONUL's Seven Pillars and Vitae's RDF to help researchers and their institutions to effectively plan data management skills development and training. Working with the *JISC 04/10: Managing Research Data programme: Promoting discipline-focused research data management skills*[35] projects, DaMSSI tested the effectiveness of the Seven Pillars Model and RDF for consistently mapping and describing data management skills and skills development paths in UK HEI postgraduate courses.

However, none of the efforts above goes far enough in defining the specific levels of skills, knowledge and understanding needed for particular roles. A prerequisite here is to reach some level of consensus on the different roles needed for effective data management. In the case of library and information professionals this means identifying the various positions or roles within libraries that should be involved in data management and curation, which could include, for example, institutional repository managers, cataloguers/metadata specialists, information literacy coordinators, and reference/subject/liaison librarians, rather than simply talking about "data librarians." We need to synthesize the work done to date, progress to more comprehensive and specific competency frameworks, then test the results in the field with practitioners in relevant roles.

[35] See JISC 04/10: Managing Research Data programme: Promoting discipline-focused research data management skills: http://www.jisc.ac.uk/fundingopportunities/funding_calls/2010/03/410dataskills.aspx (last accessed 03-26-2012).

Improving Knowledge Transfer

There are many means of knowledge transfer, such as instruction (academic and experiential); research; exchange and dissemination of knowledge through professional networks, committee work, and publications and other methods of communication and outreach; identification and recruitment of needed expertise; and exposure to fresh perspectives (e.g., through new groups of interns). Formal knowledge transfer is currently happening almost exclusively at a training level, teaching practical skills on how to manage the simpler challenges in data management, curation, and preservation. Given the amount of work going on among the various stakeholders, it is essential that we better facilitate knowledge transfer between training providers and more importantly between disciplines. Presently, many professionals are expanding their knowledge of data curation and management through participation in funded research projects. This is often where the newest challenges are being faced and, while striving to come up with a solution, teams of experts evolve. Such partnerships and collaborations have been successfully employed by the Library of Congress, evident, for example, in the network of partners and relationships across the US and the globe that the Library has built and leveraged in DPOE and the National Digital Information Infrastructure and Preservation Program (NDIIPP) and other national digital programs. Acting collaboratively is especially important in order to traverse "new, uncharted waters," leverage diverse skill sets, and build on existing infrastructure. Coordination is also important, as are open sharing and transparency to the maximum degree possible in such endeavors as:

- Training opportunities;
- Workshops and other activities;
- A repository to share the results of and lessons learned from training; and
- Publications and other information products such as Webcasts, archived videos and survey results.

In addition to traditional methods of knowledge transfer, recently developed digital tools and applications should be harnessed, including social networking platforms.

Knowledge transfer opportunities often occur as an outgrowth of national and international meetings. There should

also be regular opportunities to convene, including via teleconferences and Web conferences. However, in-person meetings that are too distant, too expensive, and/or too long can be obstacles for participants whose organizations are below a certain size or budget. In any respect possible, developing community is an important way to help people become invested in a goal or objective and to foster open sharing.

Engaging Employers and Professional Bodies

It is clear that we need to engage more with both professional bodies and prospective employers as we define curricula and develop training courses. Without their involvement, there is little chance that data management and curation skills will be recognized in the workplace or that productive professional development opportunities will be made available.

Ultimately, professional bodies should champion the cause of data management and curation. While the field has had some innovative individuals who have led high-profile projects, more needs to be done. This challenge may reflect the need for developing uniform sets of messages, procedures, and standards that can be communicated to professional bodies and, in turn, to members of professional bodies. Professional bodies could also offer materials, tutorials, and clearinghouses on data management and curation, free of charge or for a minimal cost-recovery fee. They might also provide venues for practical discussions, task forces and working groups.

Developing Accredited Trainers, Curriculum and Assessing Outcomes

As noted above, many professional development courses are offered by short-term research projects or initiatives. There is often little incentive for a professional to attend such courses as the provider is an unknown quantity. Instead, we need to engage more with professional bodies to enlist their help in promoting existing training courses on offer and—where appropriate—to cooperate on the development of accredited data management and curation training that reflects the distinctive expertise of practitioners in their fields. However, without having a solid understanding on what competences need to be taught, any accreditation currently would seem rather random.

Within the UK, the Chartered Institute of Library and Information Professionals (CILIP) accredits both educational

programs and professional practitioners, who have to evidence their competence in relation to the CILIP *Body of Professional Knowledge*[36] to become chartered members of the institute. CILIP also plays a significant role in CPD, by offering formal revalidation of professional qualifications (which may in due course be replaced by a mandatory CPD scheme) and delivering a varied program of conferences, seminars, and workshops on professional issues through its extensive network of regional branches and special interest groups. CILIP's role in promoting and supporting the development of professional roles is exemplified by the range of external and on-site courses previously on teaching and learning, which made a significant contribution to the professionalization of the teaching role of librarians in relation to information literacy development. CILIP could potentially fulfill a similar role in promoting and supporting the more extensive involvement of library and information professionals in data management and curation. Professional bodies and potential employers may also have a role to play in developing and supporting paid internships.

Where Should We Be in Five Years?

Information Science Professionals' Perspective

Within five years, data management and curation should be regarded as mainstream activity for library and information professionals, to the extent that facilitating long-term access to data sets is accepted as part of their core business and managed alongside access to other key resources and services in the continually changing information universe of the digital world. Specialist posts (e.g. data librarians, data resources managers, or coordinators) or teams (e.g. data services teams, digital curation teams) could be used within libraries to support service development and embed data services in existing library functions, such as acquisitions, cataloguing, reference, liaison and education/training, in the same way that many libraries created specialist electronic resources posts and teams to manage the transition from the print-based to the hybrid print and electronic library, until e-resources became commonplace and were no longer regarded as a new specialist activity.

---

[36] CILIP (2004) *Body of Professional Knowledge,* London: Chartered Institute of Library and Information Professionals.

Data management and curation should similarly be included in the core curriculum of initial professional education programs in library and information science, incorporated into courses/modules/units covering subjects such as information resources, information literacy, knowledge organization, collection management, intellectual property, service development, research methods, and professional roles. In addition to the integration of data management and curation into generalist programs in librarianship and information management, we should have specialist provision with more in-depth technical content for practitioners interested in specializing in this area. It is not clear how extensive such provision should be, but it could take the form of one or more electives, maybe promoted as a specialist pathway. Similarly, it is not clear whether whole programs devoted specifically to data management and curation will be needed for library/information practitioners or for entrants to the field from other backgrounds, but these could be offered at the level of a postgraduate certificate, diploma or Masters, as both initial professional education and specialized CPD programs. Such offerings already exist (notably in the iSchools at the universities of Illinois, North Carolina and Michigan), but they need to be more widespread, especially outside the US. We should also be aiming to have an array of short courses and resources for self-paced learning available for practitioners whose initial education did not cover data management, for those who decide later to specialize in this area and for general professional updating.

Success could be judged by several different criteria, depending on the perspectives taken. A key question here is whether the goal is to make data management and curation core business for library and information professionals or to create a new professional field that could be seen as either a sub-profession within the information field or a hybrid profession located at the intersection of two or more professional disciplines (for example, library/information science, information technology/computer science and/or archives/records management) or combining information-related expertise with an academic discipline.

On the one hand, success could mean data management is seen as "business as usual" and *not* something novel or specialist; on the other hand, success could mean the establishment and recognition of a new professional career, whose maturity is evidenced by the existence of specialist positions at progressive levels of the management hierarchy; development of distinct

communities of practice with their own dedicated professional networks and associations/formal bodies; and provision of specialist training courses and educational programs that are formally accredited and lead to recognized and valued qualifications.[37]

Information Technology Professionals' Perspective

One goal would be to have an accepted dual stream education scheme that is suitable to both educate experts on a foundational basis to develop the competences to develop solutions for the new data management, curation, and preservation challenges arising, as well as to train practitioners to obtain the skills to put existing know-how into practice.

Both will need to be based on extremely solid IT competences in order to understand the complexities of entire system processes. A measure of success of both the education and training activities in digital curation will be whether the experts emerging from these programs will find wide-spread acceptance in the domains where curation is currently not even being considered as a dominant topic, namely in the IT industry developing new architectures, computing principles and systems, and the industry where massive amounts of digital information (both data objects as well as entire business processes) will need to be curated or self-curating.

True success will have been reached when the concept of curation is so embedded as a standard non-functional requirement in any IT infrastructure that it would actually no longer be considered worth mentioning specifically—as a very far-reaching vision into the future.

Research Professionals' Perspective

Success would mean that data curation and management activity was seen simply as part of good research practice—a core part of any researcher's job and not something extra. Increasing funding body and research council requirements for evidence of data management planning as part of new grant proposals has led to some limited success in raising researchers' awareness of the importance of data management and curation. However, to avoid

---

[37] Corrall, S. (2008) "The emergence of hybrid professionals: new skills, roles and career options for the information professional." In: Turner, C. (ed.) *Online Information 2008 Proceedings,* pp.67-73. London: Incisive Media.

data management planning being viewed as a "tick-box" exercise by researchers, it will be increasingly important that peer reviewers are able to effectively assess data management plans that are submitted. Organizations such as the UKDA and DCC are currently developing guidelines to help reviewers assess data management plans fairly. Without strong evidence of potential rewards or demonstrable benefits for researchers undertaking data management and curation activities, it is highly unlikely that we will see any lasting success in this area.

Longer-term success could be measured by "invisible" curation where researchers simply make use of hardware and software that are capable of producing preservation-ready data. Until that time, success will depend on clear communication between all stakeholders. Researchers will need to be aware of their data management and curation responsibilities and be able to define their specific data management and curation requirements to both information technologists and information specialists. Researchers should understand the bigger picture and be able to make effective decisions about how they manage their data early on in the curation lifecycle so that longer-term curation, preservation and reuse is more easily facilitated.

Experienced Trainers' Perspective

A number of goals can be construed from progress on addressing the risks and challenges described earlier. Generally, it is hoped that the amount, depth, breadth, and flow of learning and knowledge in knowledge transfer and training in data management and curation will have expanded in five years' time across disciplinary, organizational, and national boundaries. Specifically, it is also hoped that:

1. There will be more data management and curation related internship programs offered. The concept of data management and curation training, including through internships, will have become more widespread and accepted and that data management and curation will be considered from the start in projects, grant designs, etc. Sustainability in training programs and employment will have become a reality, with budgets for opportunities that last for more than a year or two. The research base on internships will also have grown.
2. Metrics to evaluate data management and curation training and internships will have been further identified, developed, and utilized, with surveys and other data collection tools, to

study the alignment of and successes in educational and other knowledge-transfer efforts. Examples could include assessments of whether a given program or project raised the number of people skilled in aspects of data management and curation (data that can also be segmented in various ways, such as skill levels), student employment patterns, effectiveness ratings, costs over time, the number of cultural heritage institutions that have participated in data management and curation activities, and return on investment.

3. More standards, best practices, guidelines, and tools will have become available in data management and curation and internships in general. Expertise and collaboration, including on best practices, will be utilized not only across cultural-heritage fields and institutions, but from other fields (such as technology and business) where there is mutual interest and benefit. In addition to metrics, milestones will be important to incorporate into program designs, and reaching them will be key indicators of success. A consistent theme that underlies this paper is to start small, hopefully achieve some small successes, and from there build a cycle that will grow.

Recommended Areas for Alignment

Several actions could be progressed in the short term to address some of the challenges we are facing and to exploit the opportunities described above. Each of these recommendations depends upon cooperation—between disciplines, industry and at the international level—to foster any real and sustainable change in practice.

Develop Accredited Curriculum, Providers, and Metrics

As noted above, there are a number of continuing education courses incorporating aspects of data management, curation, and preservation currently on offer. Recent surveys have been carried out by DPOE[38] and DigCurV[39] to identify the number and range of training courses available across the US, Canada and EU. However, there is as yet no means of benchmarking these courses or their content. As such, it can be difficult to know who should be

[38] DPOE Needs Assessment Survey, 2010, http://www.digitalpreservation.gov/education/documents/DPOENeedsAssessmentSurveyExecutiveSummary.pdf (last accessed 05-21-2012).

[39] DigCurV project training registry, http://www.digcur-education.org/eng/Training-opportunities (last accessed 05-21-2012).

attending these courses for maximum benefit and exactly what participants will be able to do in a practical sense upon completion. We need some way to classify training courses and to illustrate clear course objectives and outcomes for prospective participants.

The RIN Information Handling Working Group has developed a draft set of criteria that could enable course providers to self-certify and quality check their courses and help to address some of the challenges listed above.[40] The draft criteria includes elements drawn from teaching and learning resources criteria devised by other bodies including Vitae,[41] Jorum,[42] CILIP,[43] HEA,[44] and DELILA.[45] While the criteria are intended to assist with self-certification in the short-term, there is longer-term potential for an external body to use the criteria as a means of formal certification of training courses.

Key recommendation:

Foster cooperation between DPOE, DigCurV and RIN Information Handling Working Group to test the draft criteria using real-life courses identified via the training surveys. There may be potential for Knowledge Exchange partners to liaise with training providers at the EU level to help test and refine the criteria. An international workshop led by RIN and the DCC to bring together training providers to review and test the criteria would be a possibility in 2012.

Address Supply and Demand

As demand for data management, curation and preservation training increases amongst all stakeholders it will be vital that there is an adequate pool of qualified trainers capable of delivering high quality tuition. DigitalPreservationEurope (DPE) developed a

---

[40] Ongoing RIN work on defining draft "Criteria for describing and reviewing good practice in information literacy training" is being led by Stephane Goldstein, RIN

[41] See Vitae Database of Practice: http://www.vitae.ac.uk/policy-practice/34837/Database-of-practice.html (last accessed 03-26-2012).

[42] See Jorum Learning and Teaching Competition: http://community.jorum.ac.uk/view.php?id=35 (last accessed 03-26-2012).

[43] CILIP CSG Information Literacy Group, Information Literacy Practitioner of the Year http://www.informationliteracy.org.uk/2010/12/csg-information-literacy-group-information-literacy-practitioner-of-the-year-nominations-sought/ (last accessed 03-26-2012).

[44] HEA evaluation of commercial online tutorial packages.

[45] DELILA criteria for evaluating information literacy and digital literacy open educational resources (OERs); these are drawn heavily from the original version of the above RIN criteria.

registry of trainers[46] to help identify individuals capable of contributing to and/or delivering data management, curation, and preservation training. While the list of experts is extensive and spans the globe, it is important to note that most of the individuals on the list are not full-time trainers. As such, there is a limit to the amount of training that they can realistically deliver. To avoid demand outstripping supply, we need to train up professional trainers, institutional support staff, and practitioners to deliver the courses where appropriate. The recent DPOE Baseline Workshop sponsored by the Library of Congress' Digital Preservation Outreach and Education (DPOE) Program aimed to develop a cohort of trainers capable of delivering curation and preservation training. Graduates of this pilot workshop were trained in six key aspects of digital preservation taught by leading experts in the field. A key component of the workshop was to guide the participants in developing and presenting their own workshops, which they will subsequently run in their own regions by the end of 2012. There is great potential for applying this approach to data management, curation and preservation in the UK, Europe and indeed worldwide.

Key recommendation:

Current training providers should evaluate the DPOE Baseline workshop approach and consider cooperating with DPOE to roll out this approach in other countries. The DCC sent a member of staff to take part in the pilot workshop as an observer and, as a result, is looking to work with DPOE to take this approach forward in the UK. The DCC will share details of the DPOE approach with fellow members of the RIN Information Handling Working Group and the Knowledge Exchange to see if there is potential for greater join-up at the EU level to run follow-on workshops in cooperation with DPOE. An initial meeting between DPOE, DCC and DigCurV to take this forward took place in October 2011 and follow-up meetings are planned.

Engage with Employers and Professional Bodies

Students need to know that participation in data management, curation, and preservation related education and training programs

[46] DigitalPreservationEurope (DPE) registry of trainers: http://www.digitalpreservationeurope.eu/registries/trainers/ (last accessed 03-26-2012).

will help them to become graduates with the actionable skills that employers seek. Most training providers provide a list of learning outcomes as part of their course descriptions but little has been done so far to actually assess whether those who participate actually leave with the ability to fulfill those learning objectives. In addition to designing mechanisms to test knowledge, it requires time and effort on the part of trainers to assess coursework and examination results. As noted above, this may be problematic due to the fact that most data management and curation trainers have other responsibilities in addition to providing training. It can also be difficult to assess data management, curation, and preservation skills in the short term. A grasp of the key concepts may only emerge as students return to the workplace and start to implement what they have learned. Networks of trainers could possibly be set up to provide ongoing feedback in a distributed fashion. But, again securing trainers' time may be a challenge. Another option might be to emulate the approach taken by DigCCurr for its professional institute.[47] Students of the professional institute are reconvened after six-to-twelve months to share how they have implemented what they learned during the course in their workplace. This approach facilitates longer-term assessment of participants' skills and places fewer ongoing demands on trainers' time.

We must also seek to engage with employers and professional bodies to act as reviewers for current training offerings and associated learning objectives and either endorse these skills or identify gaps that need to be addressed. There are a number of current initiatives looking to engage with industry and professional bodies at the moment. For instance, DaMSSI has developed a series of career profiles to illustrate how data management and curation relates to the day-to-day activities for a small number of professions. These profiles may be of value in engaging with professional bodies and industry as they demonstrate in a tangible way why data management and curation skills are important. These profiles are a great starting point but we need to develop a larger pool of profiles for a greater range of professions. The EU-funded TIMBUS project has succeeded in engaging industry as core partners in their FP7 project. Training will be a key component of this project's work and the approach adopted by TIMBUS may be a useful model for others seeking to engage with industry in the development and delivery of training.

[47] See DigCCurr Institute: http://www.ils.unc.edu/digccurr/institute.html (last accessed 03-26-2012).

APARSEN, another FP7 funded project, is aiming to develop and deliver certified training. The results of both of these projects will be of interest as they progress over the coming years.

Key recommendations:

Current training providers should review their methods of assessing participants' knowledge and skills. In particular, providers should review the DigCCurr Professional Institute model.

Training providers may wish to develop and contribute to the DCC and RIN's collection of career profiles[48] using the DaMSSI template. The profiles help to highlight the baseline data management and curation skills that professionals in various disciplines need to carry out their daily work. These profiles may also serve as useful marketing tools for attracting prospective students and could be valuable for engaging with professional bodies and industry.

Group projects could be a useful way to assess skill levels. One possible exercise would be to have students collaborate on developing a data management plan for a sample data set resulting from a fictional project. This would work particularly well for courses that aim to attract participants from multiple disciplines, as it would provide an opportunity to hone communication skills and develop a shared solution to a specific problem. Another potential means of testing skills would be to have students develop experimental strategies that can be tested in the Planets Testbed and/or Plato tool. Plato allows users to measure the effectiveness of tools to preserve at-risk objects while the Testbed provides a controlled environment to carry out preservation experiments.[49] The EC-funded Planets project came to an end in 2010 but the tools are being maintained by the Open Planets Foundation (OPF). These approaches would be suitable for both professional development training and formal education courses and could be piloted by a number of the education ANADP panel members. The DCC will aim to pilot an assessed data management planning exercise as part

[48] DaMSSI career profiles collection, http://www.dcc.ac.uk/training/data-management-courses-and-training/career-profiles (last accessed 05-21-2012).
[49] See Planets project testbed: http://www.planets-project.eu/software/ (last accessed 03-26-2012).

of its DC101 training course by the end of 2012. iSchools may be an excellent place to pilot student group projects in a formal education setting.

Improve Cooperation in Defining Skill-sets

Recent government recommendations in the UK state that HEIs should be explicit about graduates' career prospects for all courses they offer. This is something that training providers should aspire to as well. There are a number of data management and curation related courses being offered around the globe. The DPOE course calendar[50] and Digital Curation Exchange (DCE) registry list just a few. However, without agreement on how to describe courses and their objectives it is very difficult for prospective students to be able to assess which courses are right for them—both with respect to their immediate needs and also to allow them to hone their data management and curation skills over their entire careers. We need to develop a coherent way to classify education and training options to facilitate effective comparison of offerings and to enable professional development planning.

Key recommendation:

One solution might be to make use of the existing DPOE pyramid, which classifies skills into three broad categories: executive, managerial, and practical. This approach could provide a logical framework to describe courses with minimal effort on the part of course providers and potentially great benefit for students. The DPOE pyramid also lends itself to the description of course materials for those wishing to undertake self-directed learning. For instance, there could be potential to retrospectively apply the DPOE pyramid classifications to materials deposited into JORUM and the DPE registry of training materials[51] to ease discovery by prospective students. We might also wish to consider making use of the DCC's curation lifecycle model as a means of describing specific data management and curation actions and roles. The information handling aspects of Vitae's RDF may offer a valuable progression map for career development. The

[50] See DPOE training calendar: http://www.digitalpreservation.gov/education/courses/index.html (last accessed 03-26-2012).

[51] See DigitalPreservationEurope (DPE) registry of training materials: http://www.digitalpreservationeurope.eu/registries/materials/ (last accessed 03-26-2012).

results of JISC and RIN funded-DaMSSI project may offer valuable insights into the potential value of the RDF for professional development planning. The EC-funded DigCurV project is currently undertaking course profiling work for EU data management, curation, and preservation courses. Current discussions between DPOE, DCC, and DigCurV staff may result in some pilot testing of the DPOE pyramid classifications on a corpora of EU courses. Other training providers should also consider the DPOE pyramid as a means of contextualizing course offerings. Ongoing discussions between RIN, DCC and DigCurV may also result in further testing of the RDF.

Provide Hands-On Experience

There is no substitute for hands-on, practical experience. In an ideal world, we would see curation and preservation professionals emerging from something akin to a teaching hospital. Internships and student placements are another great way to boost practical skills. These exchanges are also effective for feeding employers' needs back into course design. However as noted above, a number of elements need to be built into internships and placements to ensure that they are valuable for both the host and the participant. Potential hosts and interns/students often struggle to adequately pin down what it is they are aiming to get out of the experience. As a result, many internships and placements fail to live up to either party's expectations. Host institutions that do not get interns/students with the right skills for their particular needs may be reluctant to engage in future exchanges. Similarly, we do not want to send interns/students to host institutions where their skills will not be put to best use. Success depends upon well-defined work with clear expectations—for both parties—of what will result from the experience.

Finding a raft of suitable host institutions and candidates locally can be tough. In most cases interns and students will need to consider carrying out their placements in another city or even another country. While many students are keen to carry out work experience in another country there are often linguistic, financial, and legal barriers that limit the possibilities. Regional, national, and international structures to facilitate internships and exchanges would be beneficial to both host institutions and the candidates. It would provide access to a greater pool of host institutions and suitable candidates and enable more granular matching of students/interns' skills to hosts' needs.

Key recommendation:

DPOE has established a rigorous approach to its internship program. Hosts and candidates are carefully matched to ensure that maximum benefit is achieved for both parties. The proposed work is clearly described and concrete objectives for both parties are clearly spelled out. Those aiming to offer data management, curation and preservation related placements should review the DPOE approach and consider implementing a similarly robust approach. While DPOE hosts and candidates span the US, it would be beneficial to extend this pool internationally. The authors of this essay are keen to explore the potential of extending the DPOE approach to include European partners.

Conclusions

The risks associated with a lack of alignment between disciplines and nations in developing and delivering data management, curation, and preservation education and training are serious. Numerous challenges hinder our efforts to mitigate these risks. However, there are concrete actions that could be undertaken in the short to mid-term to improve the overall outlook. There are some degrees of overlap and some dependencies in the authors' list of recommendations. Agreement in the very short term on what practical actions should be prioritized and taken forward is needed. Several of the projects and initiatives mentioned in the recommendations section are already undertaking work in key areas and could be viewed as catalysts for action.

If the recommendations cited by the authors are taken forward collectively, we should—over the next five years—be able to make good progress in:

- Describing and comparing data management courses across disciplines and match skills across data curation lifecycle and the various roles;
- Communicating data management and curation requirements and activities across disciplines;
- Making use of established frameworks to help identify progression paths for skills development in a range of disciplines;

- Assessing and benchmarking data management, curation, and preservation skills in both recent graduates and professionals; and
- Engaging with professional bodies to endorse and accredit data management and curation skills.

Data management, curation, and preservation roles and associated skill sets will change over time. Improvements to infrastructure may eventually automate and effectively shield management, curation, and preservation processes from the majority of stakeholders. However, until that point in time we need professionals in all disciplines who are trained to undertake specific management and curation actions. These professionals should also be able to communicate effectively with other stakeholders in the lifecycle. However, we must always bear in mind that mindsets are as important as skill sets. Accordingly, we must endeavor to include elements of critical thinking and problem solving in education and training courses for all disciplines along with more practical data management and curation skills.

References

ALA. (2009). *Core Competences of Librarianship,* Washington, DC: American Library Association. http://www.ala.org/educationcareers/sites/ala.org.educationcareers/files/content/careers/corecomp/corecompetences/finalcorecompstat09.pdf (last accessed 05-21-2012).

ALIA. (2005). *The Library and Information Sector: Core Knowledge Skills and Attributes,* rev. ed. Deakin: Australian Library and Information Association. http://www.alia.org.au/policies/core.knowledge.html (last accessed 05-21-2012).

APARSEN project, http://www.alliancepermanentaccess.org/index.php/aparsen/ (last accessed 05-21-2012).

ASIS&T. (2001). *ASIST Educational Guidelines,* Silver Spring, MD: American Society for Information Science and Technology. http://www.asis.org/Board/educational_guidelines.html (last accessed 05-21-2012).

CILIP. (2004). *Body of Professional Knowledge,* London: Chartered Institute of Library and Information Professionals. http://www.cilip.org.uk/sitecollectiondocuments/PDFs/qualificationschartership/BPK.pdf (last accessed 05-21-2012).

CILIP. (n.d.). *CILIP Accreditation: The Body of Professional Knowledge, A Guide for Course Designers,* London: Chartered Institute of Library and Information Professionals. http://www.cilip.org.uk/jobs-careers/qualifications/accreditation/bpk/Pages/default.aspx (last accessed 05-21-2012).

Corrall, S. (2008). "The emergence of hybrid professionals: new skills, roles and career options for the information professional." In: Turner, C. (ed.) *Online Information 2008 Proceedings,* pp. 67-73. London: Incisive Media.

DigCCurr Institute: http://ils.unc.edu/digccurr/institute.html (last accessed 03-26-2012).

DigCCurr II: Extending an International Digital Curation Curriculum to Doctoral Students and Practitioners http://www.ils.unc.edu/digccurr/aboutII.html (last accessed 03-26-2012).

DigCurV project: http://www.digcur-education.org (last accessed 03-26-2012).

Digital Curation Centre (DCC). (n.d.). Digital Curation Lifecycle Model. http://www.dcc.ac.uk/resources/curation-lifecycle-model (last accessed 05-21-2012).

Digital Curation Centre DC101 training materials: http://www.dcc.ac.uk/training/digital-curation-101 (last accessed 03-26-2012).

Digital Futures: http://www.digitalconsultancy.net/digifutures/ (last accessed 03-26-2012).

Digital Preservation Management (DPM) Workshops: http://www.dpworkshop.org (last accessed 03-26-2012).

Digital Preservation Outreach and Education (DPOE): http://www.digitalpreservation.gov/education/ (last accessed 03-26-2012).

Digital Preservation Training Programme (DPTP):

http://www.dptp.org/ (last accessed 03-26-2012).

Gabridge, T. (2009). "The last mile: liaison roles in curating science and engineering research data," *Research Libraries Issues,* 265, pp. 15-21. http://www.arl.org/bm~doc/rli-265-gabridge.pdf (last accessed 05-21-2012).

Garritano, J.R. and Carlson, J.R. (2009). "A subject librarian's guide to collaborating on e-science projects," *Issues in Science and Technology Librarianship,* 57. http://www.istl.org/09-spring/refereed2.html#15 (last accessed 05-21-2012).

Hank, C. and Davidson, J. (2009). "International Data curation Education Action (IDEA) Working Group: a report from the second workshop of the IDEA," *D-Lib Magazine,* 15 (3/4). http://www.dlib.org/dlib/march09/hank/03hank.html (last accessed 05-21-2012).

Henty, M. (2008). "Developing the capability and skills to support e-research," *Ariadne,* 55. http://www.ariadne.ac.uk/issue55/henty/ (last accessed 05-21-2012).

High Level Expert Group on Scientific Data. (2010). *Riding the Wave*: *How Europe can gain from the rising tide of scientific data* report. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf (last accessed 05-21-2012).

Howe, D. *et al.* (2008). "Big data: the future of biocuration," *Nature,* 455 (7209), pp. 47-50. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2819144/ (last accessed 05-21-2012).

IFLA (2003). *Guidelines for Professional Library/Information Educational Programs,* 3rd rev. ed. The Hague: International Federation of Library Associations and Institutions. http://www.ifla.org/en/publications/guidelines-for-professional-libraryinformation-educational-programs-2000 (last accessed 05-21-2012).

Information Management and Preservation (IMP) programme at Humanities Advanced Technology and Information

Institute at the University of Glasgow. http://www.gla.ac.uk/postgraduate/taught/informationmanagementpreservationdigitalarchivesrecordsmanagement/ (last accessed 05-21-2012).

JORUM: http://www.jorum.ac.uk (last accessed 03-26-2012).

Knowledge Exchange: http://www.knowledge-exchange.info/ (last accessed 03-26-2012).

Lanier D, Henderson C. (1999). "Library residencies and internships as indicators of success: evidence from three programs." *Bulletin of the Medical Libraries Association*. 87(2):192-199.

Lewis, M. (2010). "Libraries and the management of research data," in McKnight, S. (ed.), *Envisioning Future Academic Library Services: Initiatives, Ideas and Challenges,* pp. 145-168, Facet, London.

Library of Congress. (2010). Digital Preservation Outreach and Education (DPOE) Training Needs Assessment Survey: Executive Summary. http://digitalpreservation.gov/education/documents/DPOENeedsAssessmentSurveyExecutiveSummary.pdf (last accessed 05-21-2012).

Macdonald, S. and Martinez, L. (2005). "Supporting local data users in the UK academic community," *Ariadne,* 44. http://www.ariadne.ac.uk/issue44/martinez/ (last accessed 05-21-2012).

NSF. (n.d.) *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540_5.pdf (last accessed 05-21-2012).

Open Planets Foundation: http://www.openplanetsfoundation.org/ (last accessed 03-26-2012).

Pryor, G. and Donnelly, M. (2009). "Skilling up to do data: whose role, whose responsibility, whose career?" *International Journal of Digital Curation,* 4 (2),158-170. http://www.ijdc.net/index.php/ijdc/article/view/126 (last accessed 05-21-2012).

Ray, J. (2009). "Sharks, digital curation, and the education of information professionals," *Museum Management and*

Curatorship, 24 (94), 357-368.

SCONUL (2011a). *The Seven Pillars of Information Literacy: A Research Lens for Higher Education,* London: Society of College, National and University Libraries, Working Group on Information Literacy. http://www.sconul.ac.uk/groups/information_literacy/publications/researchlens.pdf (last accessed 05-21-2012).

SCONUL (2011b). *The Seven Pillars of Information Literacy: Core Model for Higher Education,* London: Society of College, National and University Libraries, Working Group on Information Literacy. http://www.sconul.ac.uk/groups/information_literacy/publications/coremodel.pdf (last accessed 05-21-2012).

Sides CH, Mrvica A. (2007). *Internships: Theory and Practice.* Amityville, NY: Baywood Publishing.

Swan, A. and Brown, S. (2008). *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs,* Report to the JISC, Truro: Key Perspectives. http://www.jisc.ac.uk/publications/documents/dataskillscareersfinalreport.aspx (last accessed 05-21-2012).

TIMBUS project, http://timbusproject.net/ (last accessed 05-21-2012).

UK Data Archive: http://www.data-archive.ac.uk/ (last accessed 03-26-2012).

Vitae (2010). *Researcher Development Framework*. Cambridge: Vitae. http://www.vitae.ac.uk/rdf (accessed 05-21-2012).

Witt, M. (2008). "Institutional repositories and research data curation in a distributed environment," *Library Trends,* 57 (2), 191-201.